Towards Modernizing Rice Yield Statistics in Indonesia: Linking Crop-Cutting Survey with Remote Sensing Data[1]

Octavia Rizky Prasetyo
Statistics Indonesia (BPS), Jakarta, Indonesia – octavia.rizky@bps.go.id
Bagus Setyawan Wijaya
School of Electrical Engineering and Informatics, Institut Teknologi Bandung
Widyo Pura Buana
Statistics Indonesia (BPS), Jakarta, Indonesia – wiwid@bps.go.id
Sugi Hariyanto
Statistics Indonesia (BPS), Jakarta, Indonesia – sugih@bps.go.id

Abstract

Rice is an essential crop in Indonesia, making accurate and timely rice production statistics critical for agricultural policymaking. Currently, rice yield estimates from Statistics Indonesia (BPS) rely on Crop-Cutting Survey (CCS). Although reliable, CCS is labor-intensive, costly, and subject to delays. Recent advances in Earth Observation (EO), particularly Sentinel-2, offer opportunities to improve efficiency and timeliness of rice yield estimation. This study demonstrates a workflow integrating CCS data with Sentinel-2 imagery and climatic covariates to predict rice yields in Indramayu Regency, a key rice-producing area in West Java, using 2023 CCS samples. Analysis focused on *Subround* III (September–December) to reduce seasonal variability. Sentinel-2 data were processed to derive vegetation indices within spatial buffers around CCS locations, and Random Forest models were evaluated across harvest months. Results reveal seasonal growth differences and show that models combining peak NDVI and climatic covariates outperformed vegetation indices alone. The study demonstrates a feasible approach for integrating EO into official yield estimation, highlighting both potential and operational challenges.

Keywords: Crop-cutting survey, rice yield, agricultural statistics, remote sensing, Sentinel-2

1. Introduction

Rice is an essential crop in Indonesia, making accurate and timely rice production statistics critical for agricultural policymaking. Official estimates of rice production are currently generated by Statistics Indonesia (BPS) using a combination of two survyes: Area Sampling Frame (ASF) survey for harvested area and Crop-Cutting Survey (CCS) for yield measurement [1,2]. The ASF survey enables monthly harvested area updates through enumerator visits to sample segments, where rice growth phases are identified and geotagged locations recorded [1]. CCS, conducted annually in three four-monthly periods (*subround*), provides plot-level yield measurement from

[1] The text and materials presented in this article are free from any copyright violations.

Acknowledgement: *The authors would like to thank Dr. Simon Collings and Dr. Joanne Chia for their technical guidance and support in understanding methodological aspects of this study.*

2.5 m x 2.5 m rice plots through standardized field procedures [2], offering statistically robust and nationally comparable estimates.

Despite its reliability, CCS remains labor-intensive, costly, and logistically demanding. With around 90 thousand plots are targeted annually, the survey requires substantial human resources. Enumerators frequently face challenges such as difficult access to remote areas, early harvesting, and equipment shortages. Moreover, yield estimates are typically available around two months after each four-monthly data collection, limiting their usefulness for near real-time policy decisions-making.

At the same time, Earth Observation (EO) data has become increasingly accessible, with medium-resolution satellite data such as Sentinel-2, providing frequent, multispectral observations suitable for rice growth monitoring and yield estimation [3,4]. EO-based indicators, including vegetation indices such as the Normalized Difference Vegetation Index (NDVI), have been widely explored for crop monitoring and yield prediction [5-8]. However, most studies focus on methodological advance rather than practical integration with existing official survey systems.

This study aims to propose and demonstrate a workflow linking existing ground-truth data from CCS with Sentinel-2 imagery and climatic covariates to predict rice yield. Using a case study from Indramayu Regency, West Java, we assess the feasibility of complementing conventional survey-based yield estimation with with EO data. The study emphasizes methodological transparency, operational constraints, and lessons relevant for the modernization of agricultural statistics.

2. Study Area

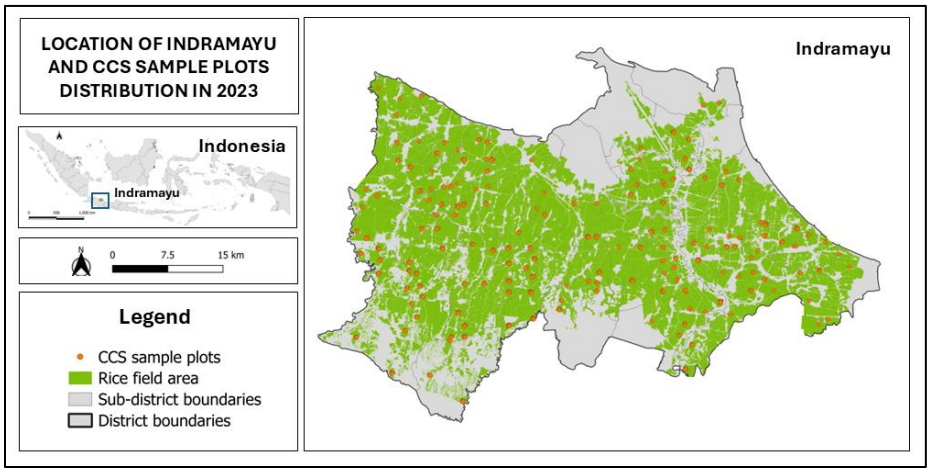

Figure 1. Location map, rice field area, and sample plots in Indramayu, 2023

The study area is Indramayu Regency (Figure 1), West Java's largest rice-producing regency. The regency lies along the northern coast of Java Island (107º52'-108º52' E, 6º15' -6º40' S). Indramayu is widely recognized for its agricultural landscape, particularly rice cultivation, with rice fields covering approximately 122 thousand hectares, representing around 60% of its land area. With its wide rice field area, Indramayu has made significant contributions to rice production, yielding 1.42 million tonnes of milled dry rice in 2023 [9]. Agricultural landholdings in the region are typically small, averaging less than one hectare per household [10].

Climatic conditions are strongly influenced by the monsoon system [11], resulting in distinct wet and dry seasons that affecting planting schedules, crop growth duration, and availability of cloud-

free satellite observations. The cropping calendar is generally divided into one wet-season followed by two dry-seasons. Rice cultivation is predominantly wetland rice [12], with a typical growth cycle of around 3–4 months (100–120 days).

3. Materials and Method

Figure 2 illustrates the schematic workflow integrating CCS ground-truth yield data with the EO-based indicators for rice yield prediction. The dataset used in this study includes rice yield measurements from CCS, multi-temporal Sentinel-2 imagery, and climatic variables.

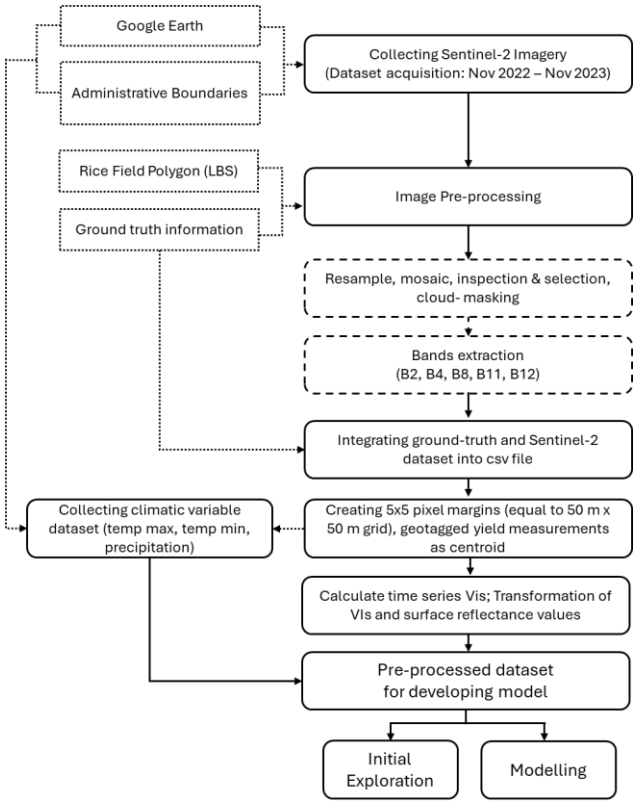

Figure 2. Schematic workflow applied in integrating CCS and EO-based data for yield prediction

3.1 Data Sources

3.1.1 Crop-Cuting Survey (CCS) Dataset

In this study, CCS provides ground truth yield data to be linked with EO indicators. CCS is conducted by BPS annually in four-monthly periods (*Subround*) covering January-April (*Subround* I), May-August (*Subround* II), and September-December (*Subround* III). CCS uses ASF survey results as its sampling frame, targeting subsegments (100 m x 100 m size) with crop stages expected to reach harvest in the subsequent months, such as early vegetative, final vegetative, generative, and land preparation [2]. Selected subsegment are assigned to enumerators around two months prior to data collection period, with parcel locations guided by ASF geotagged coordinates.

Enumerators are deployed across Indonesia's region to directly measure rice yield using standardized crop-cutting equipment within randomly selected plots measuring 2.5 m x 2.5 m, located inside selected rice parcels. In addition, enumerators collect additional information on crops by farmer interviews. While CCS provides statistically robust estimates, its design was not

intended for satellite integration. Key metadata like exact sowing dates are not systematically recorded. For this study, we focused on those CCS observations lying within the 2023 rice field area boundaries, providing a total of 507 observations in the region (Figure 1).

3.1.2 Sentinel-2 Satellite Imagery

Earth observation data used in this study derived from Sentinel-2 satellite data. Sentinel-2 provides multispectral imagery at 10 m to 60 m resolution with a five-days revisit frequency [13]. Sentinel-2 mission includes 13 spectral bands spanning visible, near-infrared, and shortwave infrared of electromagnetic spectrum. With its relatively fine spatial resolution, rich spectral information, and frequent revisit cycle, Sentinel-2 offers strong potential for agricultural applications, including crop monitoring and mapping [3,4].

Most remote sensing-based crop monitoring studies utilize vegetation indices derived from combinations of spectral bands to enhance sensitivity to vegetation characteristics [14]. Among these indices, NDVI is the most widely used indicator in crop phenological monitoring and yield-related studies. NDVI exploits the contrast between red and near-infrared reflectance to provide a proxy measure of vegetation greenness and photosynthetic activity, making it suitable for characterizing crop growth dynamics.

Sentinel-2 images used in this study were obtained from Google Earth Engine (GEE) platform, covering 1 November 2022 to 30 November 2023 period, using Google Colab platform with Python programming. Preprocessing included downloading and resampling raw bands to a 10-meter reference grid. The complete images series were initially used to explore rice phenological patterns prior to modelling.

3.1.3 Climatic Data

Crop yields, including rice, are sensitive to climatic variability, making climatic variables important for yield prediction [15-18]. Monthly precipitation and temperature data corresponding to CCS observations were obtained from TerraClimate datasets via GEE. These variables complement satellite-derived indicators by capturing inter-monthly variability in growing conditions that may not be fully reflected in vegetation indices alone.

3.2 Data Pre-processing and Feature Construction

3.2.1 Sentinel-2 Image Preparation, Cloud Masking, and Vegetation Index Derivation

All Sentinel-2 images were resampled to a 10-meter reference grid using nearest-neighbor interpolation [17]. Cloud contamination was addressed through manual cloud masking in QGIS, excluding cloud-covered areas within rice field polygon to ensure only cloud-free pixels were retained. Surface reflectance values were extracted from Sentinel-2 spectral bands commonly used in crop monitoring (B2, B4, B8, B11, and B12). From these bands, vegetation indices were derived to characterize crop development over time, including NDVI, SAVI [19,20], and RGVI [21,22] (see Table 2).

3.2.2 Spatial Matching and Pixel Window Creation

To ensure spatial consistency between CCS yield observations and Sentinel-2 imagery, each yield point was first aligned to the nearest Sentinel-2 pixel centroid using Euclidean distance. At 10 m resolution a single Sentinel-2 pixel represents an area substantially larger than the 2.5 x 2.5 m CCS crop cutting plot. Therefore, direct use of a single pixel may be sensitive to geolocation error and mixed pixel effects. To mitigate these issues, a 5 x 5 pixel window (approximately 50 m x 50 m)

was generated around the aligned centroid, bounded by rice field boundaries (Figure 3). This configuration balances spatial specificity and noise reduction, where smaller window risk positional inaccuracies while larger windows may include neighboring parcels with different growth stages. Vegetation index values were aggregated by computing the mean across the 25 pixels to construct time series for each CCS observation.

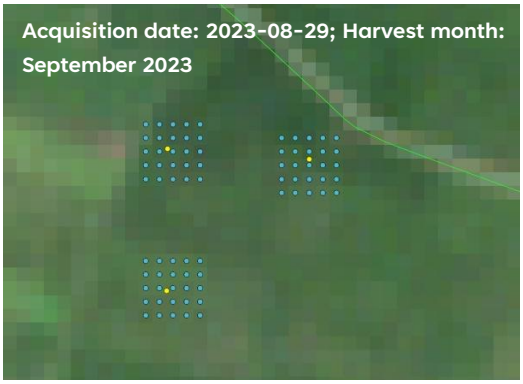

Figure 3. Example of 5 x 5 pixel window creation

3.2.3 Temporal Function Fitting and Gap Filling

Satellite-derived VIs often contains temporal gaps and noise from cloud cover and irregular acquisition. Smoothing and gap-filling techniques are essential to address the influences of clouds in VI time series [23]. A non-linear smoothing and function fitting approach was applied to NDVI time series to reconstruct continuous crop growth trajectories. For each CCS observation, a cubic smoothing spline was fitted to NDVI time series using the *smooth.spline* function in R, interpolating observed values with piecewise polynomial functions to produce smooth, continuous curves while reducing noise [25]. The smoothing parameter (*spar*) controlled the trade-off between how closely the curve follows the observed NDVI values and the smoothness of the reconstructed curve. This approach accommodates irregularly spaced acquisition dates and implicitly fills missing observations caused by cloud contamination.

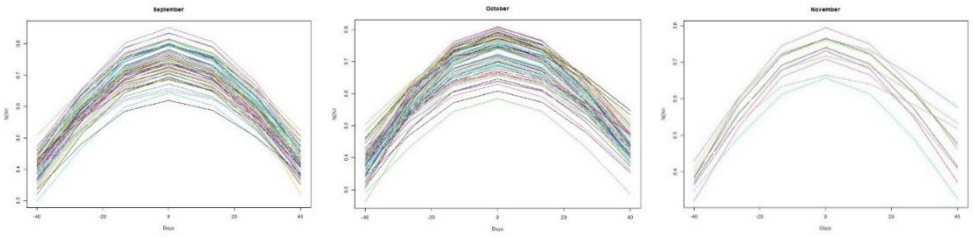

Figure 4. Phenological profiles of rice using smoothed NDVI

Since CCS record only harvest month without sowing or exact harvest dates, temporal reconstruction used a four-month window corresponding to each harvest month, defined based on typical 100–120 days rice growth cycle in Indramayu. For example, observations harvested in August 2023, imagery from May–August 2023 was included to capture the full growth cycle preceding harvest. Following curve fitting, the *optimize()* function identified peak NDVI values. To ensure phenological comparability across plots with different planting dates, predicted NDVI series were temporally aligned relative to the identified peak (Figure 4). To reduce the influence of noisy or uninformative signals, an outlier filtering step was applied. Outlier filtering removed observations with low temporal variability (quantile-based threshold), which likely reflects poor

signal quality or insufficient crop development information. These procedures were consistently applied to all EO-derived predictors.

3.3 Modelling Strategy

3.3.1 Temporal Scope and Period Selection

Although Sentinel-2 imagery was initially processed for full annual cycle, variability in crop growth across harvest months, uneven sample sizes, and seasonal cloud cover introduced additional complexity. To ensure more consistent temporal alignment and data availability, modelling analysis focused on *Subround* III (September–November), excluding December due to insufficient CCS samples (Table 1). This decision was guided by three considerations.

Table 1. CCS Samples Distribution across Months in Indramayu, 2023

No	Harvest Month	Count CCS Sample	Min Yield (kg/plot)	Max Yield (kg/plot)	Average Yield (kg/plot)	Standard deviation (kg/plot)
1	February	36	2.34	5.94	4.70	0.88
2	March	120	2.17	7.12	4.94	0.94
3	April	90	3.5	7	5.16	0.81
4	May	54	1.51	6.18	4.54	0.96
5	June	4	3.52	3.98	3.64	0.23
6	July	11	2.69	4.88	3.71	0.74
7	August	20	2.9	6.64	4.98	0.77
8	**September**	76	1.87	7.46	4.80	0.96
9	**October**	84	2.47	8	5.02	1.23
10	**November**	12	3	5.46	4.30	0.77
Total		**507**	**1.51**	**8**	**4.86**	**0.99**

Notes: No harvest events recorded in January and December for rice crops during 2023 survey period.

First, exploratory analysis of NDVI time series revealed substantial variability in growth cycle patterns across harvest months. Earlier periods (e.g. February–April) exhibited longer and more heterogeneous growth trajectories, while September–November observations showed relatively consistent patterns, improving phenological comparability. Second, Subround III provided the most balanced and sufficient sample sizes within the latter half of the year (September, n = 76; October, n = 84), whereas months like June (n=4) and July (n=11) were too limited for reliable modelling. Third, cloud-free Sentinel-2 acquisitions were more frequent during this period, enabling more robust temporal smoothing and peak extraction.

Based on these factors, modelling experiments were conducted for September, October, and a combined September–October dataset. November (n=12) was retained for exploratory analysis and out-of-sample testing but excluded from standalone modelling. This approach allowed assessment of within-month predictive performance and models transferability across adjacent harvest months.

3.3.2 Modelling Approach and Predictor Variables

To evaluate predictive capability, we implemented both parametric and non-parametric approaches. Linear regression (LM) served as a conventional baseline. Random Forest (RF)

regression was selected as the primary method due to its ability in capturing nonlinear relationships, robustness to multicollinearity among spectral and climatic covariates, and its ability to reduce overfitting. CCS plot-level yield serves as the response variable. Predictor configurations (Table 2) included vegetation indices and climatic variables.

Table 2. Variables used in crop yield estimation

Type	Dataset	Definition	Source
Vegetation indices	$NDVI_t$	Normalized Different Vegetation Index time series, calculated from $$NDVI = \frac{NIR - Red}{NIR + Red}$$	Sentinel-2
	$SAVI_t$	Soil Adjusted Vegetation Index time series, calculated from $$SAVI = \frac{(NIR - Red)}{(NIR + Red + L)} \times (1 + L)$$	Sentinel-2
	$RGVI_t$	RGVI time series, calculated from $$RGVI = 1 - \left(\frac{Blue + Red}{NIR + SWIR1 + SWIR2}\right)$$	Sentinel-2
	$NDVI_{peak}$	Peak of Normalized Different Vegetation	Sentinel-2
Climatic variables	Precipitation	Average of monthly precipitation accumulation before harvest month	TerraClimate
	T_{max}	Average of monthly temperature maximum before harvest month	TerraClimate
	T_{min}	Average of monthly temperature minimum before harvest month	TerraClimate

3.3.3 Validation and Hyperparameter Tuning

Model performance was assessed using 10-fold cross-validation conducted separately within each harvest month (Figure 5). For each configuration, data were randomly partitioned into ten subsets. In each iteration, nine subsets were used for training, and the remaining subset for validation. This procedure was repeated until all subsets served for validation, with performance metrics averaged across folds.

For RF model, hyperparameter tuning was conducted within the cross-validation framework using grid-search (Table 3). Optimal parameter (*mtry*, *ntree*, and *maxnodes*) were selected based on minimizing cross-validated root mean square error (RMSE). Following optimization, the final model was retrained using the full dataset for the respective harvest month.

Table 3. Hyperparameter ranges for RF methods

Hyperparameters	Range
Number of features in each split	{2,3,4,5}
Number of trees	{100,200,350,500,1000}
Max terminal nodes	{10,20,30}

3.3.4 One-Month-Ahead Prediction Scenarios

To assess operational feasibility for near real-time yield forecasting, we designed experiments to predict yield one month ahead. Three scenarios were implemented within *Subround* III: 1) Scenario 1, train on September data, predict October yields; 2) Train on October data, predict November yields; 3) Train on combined September-October data, predict November yields. These scenarios evaluate whether relationships learned under specific seasonal conditions can generalize to adjacent harvest periods. Prediction accuracy was assessed using RMSE and mean absolute error (MAE). Observed variability in transferability provides insights into seasonal dynamics and potential limitations for operational implementation.

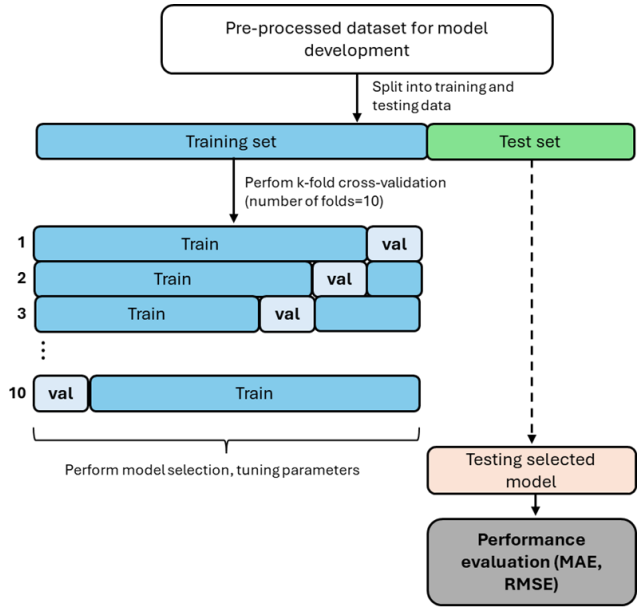

Figure 5. Cross-validation and modelling strategies

4. Results and Discussion

4.1 Phenological Dynamics Across Harvest Months

Exploratory analysis of raw NDVI time series reveals substantial variability in rice growth trajectories across harvest months. Early harvest months (e.g. February-April) exhibited longer, more heterogeneous cycles with noticeable cloud-induced data gaps, complicating peak identification. In contrast, NDVI trajectories for *Subround* III exhibited clearer, more coherent growth patterns, with higher cloud-free acquisitions density, enabling more reliable smoothing and peak extraction. These differences in temporal signal quality motivated the modelling focus on *Subround* III. Reliable temporal reconstruction requires sufficient EO observations across the growth cycle, particularly when exact sowing and harvest dates are unavailable. *Subround* III provides a suitable balance between sample size and EO data availability, allowing spline-based smoothing and peak extraction with greater confidence.

4.2 Within-Month Model Performance

Within-month predictive demonstrated that Random Forest (RF) models generally outperformed linear regression (LM), particularly when nonlinear combinations of vegetation indices and climatic variables were included (Table 4).

Table 4. Within-month cross-validation performance (10-fold)

Harvest Month	Model	LM		RF	
		MAE (kg/plot)	RMSE (kg/plot)	MAE (kg/plot)	RMSE (kg/plot)
September [Stdv= 0.96]	$NDVI_t$	0.6766	0.8880	0.6850	0.8667
	$NDVI_t$ + climatic var	0.9782	1.1765	0.6311	0.8107
	$NDVI_{peak}$ + climatic var	0.6891	0.8657	0.5661	0.7466
	$RGVI_t$	0.6837	0.8680	0.6709	0.8385
	$SAVI_t$	0.6734	0.8667	0.6851	0.8667
October [Stdv= 1.23]	$NDVI_t$	0.9745	1.1400	0.9856	1.1835
	$NDVI_t$ + climatic var	0.6991	0.8691	0.8069	1.0097
	$NDVI_{peak}$ + climatic var	0.9683	1.1591	0.7426	0.9773
	$RGVI_t$	1.0131	1.1723	0.8665	1.0614
	$SAVI_t$	0.9193	1.0944	0.9902	1.1876
September-October [Stdv=1.10]	$NDVI_t$	0.8445	1.0432	0.8203	1.0230
	$NDVI_t$ + climatic var	0.6897	0.8885	0.7352	0.9344
	$NDVI_{peak}$ + climatic var	0.8571	1.0526	0.6779	0.9022
	$RGVI_t$	0.8419	1.0248	0.8043	0.9937
	$SAVI_t$	0.8589	1.0542	0.8281	1.0282

For September, models based solely on NDVI time-series features produced moderate accuracy. Incorporating climatic variables reduced prediction error, especially under RF specification. The best-performing configuration was RF model combining peak NDVI and climatic variables, yielding the lowest MAE (0.566 kg/plot) and RMSE (0.747 kg/plot). This indicates that peak canopy development, interpreted alongside temperature and precipitation, provides stronger representation of yield variability than vegetation indices alone. October showed similar pattern with slightly higher errors, consistent with greater yield variability (SD=1.23 kg/plot). EO-only models exhibited larger errors, whereas models incorporating climatic covariates demonstrated improved performance. Again, the RF model using peak NDVI and climatic variables achieved the most favorable accuracy metrics (MAE=0.743 kg/plot, RMSE=0.977 kg/plot). The combined September and October datasets maintained stable performance, though slightly lower than single-month configurations due to increased heterogeneity from merging two harvest periods. Nonetheless, the RF model combining peak NDVI and climatic variables continued to outperform alternatives, indicating that peak-based features are robust predictors across adjacent months when supported by climatic information.

Among alternative vegetation indices, SAVI and RGVI generally produced weaker predictive performance compared to NDVI-based models, suggesting that peak NDVI captures the most relevant canopy-related signal for yield estimation. Three key findings emerge: 1) machine learning approaches, particularly RF, provide more accurate yield predictions than linear

regression; 2) climatic variables consistently enhance predictive performance; and 3) peak NDVI with climate variables emerges as the most effective predictor configuration.

4.3 One-Month-Ahead Prediction Performance

Operational implementation requires understanding whether models trained in one harvest month can generalize to subsequent months. One-month-ahead prediction experiments were conducted by training models on one month and applying them to predict yields in the following month (Table 5). One-month-ahead prediction showed lower accuracy than within-month results, indicating temporal transferability limitation. September-trained models predicting October yielded higher errors compared to within-September validation. October-trained models predicting November performed better, while combined September-October training for November prediction achieved intermediate performance.

Table 5. One-month-ahead prediction performance

Training to Test	Model	LM		RF	
		MAE (kg/plot)	RMSE (kg/plot)	MAE (kg/plot)	RMSE (kg/plot)
Sep to Oct	$NDVI_t$	1.0403	1.3093	0.9716	1.1803
	$NDVI_t$, climatic var	1.6545	1.9388	0.9392	1.1351
	$NDVI_{peak}$, climatic var	1.2255	1.3950	0.9643	1.1792
Oct to Nov	$NDVI_t$	0.9618	1.0752	0.8367	0.9961
	$NDVI_t$, climatic var	0.9722	1.2596	0.6028	0.6735
	$NDVI_{peak}$, climatic var	0.8900	0.9884	0.5656	0.6401
Sep – Oct to Nov	$NDVI_t$	0.9163	0.9999	0.8378	0.9495
	$NDVI_t$, climatic var	0.9486	1.1006	0.6236	0.7102
	$NDVI_{peak}$, climatic var	0.9366	1.0104	0.6245	0.6836

Climatic variables consistently improved transfer accuracy, suggesting they help account for inter-month environmental variability. Peak NDVI with climatic variables generally maintained better transfer than those relying solely on full time-series NDVI features, likely because peak development represents a physiologically meaningful growth milestone that is less sensitive to slight shifts in acquisition timing. However, performance degradation relative to within-month validation suggests yield and spectral relationships vary seasonally, requiring month-specific calibration or rolling model updates for operational implementation.

4.4 Implications of Integrating EO into Official Rice Yield Estimation

This study provides several insights for modernizing rice yield estimation within official statistical systems. By linking CCS data with Sentinel-2-derived indicators and climatic variables, we demonstrate a feasible workflow that complements existing survey-based approaches. First, EO-derived indicators explain substantial yield variability during specific periods, supporting their use as complementary predictive layers when rapid updates are required. Second, consistent improvement following climatic variables inclusion underscores the importance of environmental context beyond vegetation indices alone. Third, temporal transferability limitations indicate that season-specific calibration is necessary rather than adopting a single year-round model.

Importantly, this study reinforces that EO data should complement rather than replace CCS. CCS remains the benchmark for ground-truth yield measurement, providing nationally representative estimates. However, EO-derived models can help address operational challenges, including reporting delays and resource constraints, by providing early yield indications prior to full survey compilation. Several practical considerations must be acknowledged. Absence of detailed phenological metadata in CCS limits precise temporal alignment between ground observations and satellite-derived signals. Improving metadata collection within existing survey instruments could substantially enhance temporal alignment and model performance.

5. Conclusion and Recommendation

This study demonstrates a feasible workflow for integrating CCS data with Sentinel-2 imagery and climatic covariates to predict yields in Indramayu. Results show that EO-derived vegetation indices capture meaningful yield variability, particularly peak NDVI with climatic variables. RF models consistently outperformed LM, indicating that nonlinear relationships between canopy development, environmental conditions, and yield outcomes are important for explaining productivity differences. Within month validation demonstrated promising predictive capability during *Subround* III, while one-month-ahead predictions revealed temporal transferability limitations requiring season-specific calibration strategies. EO integration should complement rather than replace CCS, offering opportunities for early indication, efficiency gains while maintaining survey-based benchmark.

Although this study is limited to a single regency and one *subround*, the workflow is scalable. Future research should enhance phenological metadata collection, expand temporal scope across multiple years and *subrounds*, incorporate additional auxiliary variables, explore combined use of Sentinel-2 optical imagery and Synthetic Aperture Radar (SAR) data, and develop multi-year modelling frameworks to enhance generalization. Overall, these findings support gradual EO integration into official rice yield estimation system for more timely, adaptive, and responsive agricultural statistics production.

References

[1] BPS. *Luas Panen dan Produksi Padi di Indonesia 2024 (Hasil Kegiatan Pendataan Statistik Pertanian Tanaman Pangan Terintegrasi dengan Metode Kerangka Sampel Area).* Vol. 7. Jakarta: BPS; 2025.

[2] BPS. The Analysis of Paddy Yield in Indonesia 2024 (The Results of Crop Cutting Survey). Vol. 5. Jakarta: BPS; 2025.

[3] Soriano-González, J., Angelats, E., Martínez-Eixarch, M., Alcaraz, C. Monitoring rice crop and yield estimation with Sentinel-2 data. *Field Crops Research*. 2022;281:108507.

[4] You N, Dong J. Examining earliest identifiable timing of crops using all available Sentinel 1/2 imagery and Google Earth Engine. *ISPRS J Photogramm Remote Sens.* 2020;161:109–23.

[5] Choudhary, K., Shi, W., Dong, Y., Paringer, R. Random Forest for rice yield mapping and prediction using Sentinel-2 data with Google Earth Engine. *Advances in Space Research*. 2022;70(8):2443-2457.

[6] Islam, M. M., Matsushita, S., Noguchi, R., Ahamed, T. Development of remote sensing-based yield prediction models at the maturity stage of boro rice using parametric and nonparametric approaches. *Remote Sensing Applications*: *Society and Environment*. 2021;22:100494.

[7] Fernandez-Beltran, R., Baidar, T., Kang, J., Pla, F. Rice-yield prediction with multi-temporal sentinel-2 data and 3D CNN: A case study in Nepal. *Remote sensing*. 2021;13(7): 1391.

[8] Yanti D, Safitri I, Rusnam R, Stiyanto E. Rice productivity estimation using remote sensing method. 2022.

[9] BPS Provinsi Jawa Barat. *Luas Panen dan Produksi Padi di Provinsi Jawa Barat 2023 (Angka Tetap)*. Jawa Barat: BPS Provinsi Jawa Barat; 2024.

[10] BPS Provinsi Jawa Barat. *Hasil Survei Pertanian Antar Sensus (SUTAS) 2018 Provinsi Jawa Barat Seri-A2*. Jawa Barat: BPS Provinsi Jawa Barat; 2019.

[11] Hidayati R, Chrisendo DN. Prediction of planting date and growing period using sea surface temperature (SST) anomalies in Nino 3.4 for Indramayu District. *Agromet.* 2010;24(2):1–8.

[12] BPS Kabupaten Indramayu. *Hasil Pencacahan Lengkap Sensus Pertanian 2023 - Tahap II: Usaha Pertanian Perorangan (UTP) Tanaman Pangan Kabupaten Indramayu*. Indramayu: BPS Kabupaten Indramayu; 2024.

[13] Drusch M, Del Bello U, Carlier S, Colin O, Fernandez V, Gascon F, et al. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sens Environ.* 2012;120:25–36.

[14] Zeng L, Wardlow BD, Xiang D, Hu S, Li D. A review of vegetation phenological metrics extraction using time-series, multispectral satellite data. *Remote Sens Environ.* 2020;237:111511.

[15] Prasad AK, Chai L, Singh RP, Kafatos M. Crop yield estimation model for Iowa using remote sensing and surface parameters. *Int J Appl Earth Obs Geoinf.* 2006;8(1):26–33.

[16] Rachmawati RR, Handoko U, Purba SF, Sari UK, Saleh Y, Zahara Z, et al. Regional variability in climate–rice yield relationships across Java, Indonesia: a 30-year spatial-temporal analysis. *Int J Des Nat Ecodyn.* 2025;20(8):1815–25.

[17] Desloires J, Ienco D, Botrel A. Out-of-year corn yield prediction at field-scale using Sentinel-2 satellite imagery and machine learning methods. *Comput Electron Agric.* 2023;209:107807.

[18] Farmonov N, Amankulova K, Szatmári J, Urinov J, Narmanov Z, Nosirov J, et al. Combining PlanetScope and Sentinel-2 images with environmental data for improved wheat yield estimation. *Int J Digit Earth.* 2023;16(1):847–67.

[29] Amankulova K, Farmonov N, Omonov K, Abdurakhimova M, Mucsi L. Integrating the Sentinel-1, Sentinel-2 and topographic data into soybean yield modelling using machine learning. *Adv Space Res.* 2024;73(8):4052–66.

[20] Huete, A. R. A soil-adjusted vegetation index (SAVI). *Remote sensing of environment*, 1988;25(3):295-309.

[21] Htun AM, Shamsuzzoha M, Ahamed T. Rice yield prediction model using normalized vegetation and water indices from Sentinel-2A satellite imagery datasets. *Asia Pac J Reg Sci.* 2023;7(2):491–519.

[22] Franch B, Bautista AS, Fita D, Rubio C, Tarrazó-Serrano D, Sánchez A, et al. Within-field rice yield estimation based on Sentinel-2 satellite data. *Remote Sens.* 2021;13(20):4095.

[23] Liang S, Wang J. Compositing, smoothing, and gap-filling techniques. In: Liang S, Wang J, editors. *Advanced remote sensing*. 2nd ed. Amsterdam: Academic Press; 2020. p. 107–30.

[24] Wolberg G, Alfy I. Monotonic cubic spline interpolation. In: Proceedings of Computer Graphics International 1999; 1999. p. 188–95.